A Dual Tri-Path CNN System for Brain Tumor Segmentation

Juncheng Tong and Chunyan Wang

Department of Electrical and Computer Engineering, Concordia University,

1455 De Maisonneuve Blvd. W. Montreal, Quebec, Canada, H3G 1M8

Corresponding author: Chunyan Wang (chunyan@ece.concordia.ca).

Abstract The research on developing CNN-based fully-automated brain-tumor-segmentation systems has been progressing rapidly. For the systems to be applicable in practice, a good processing quality and reliability are necessary. Moreover, as the parameters in a CNN are determined by training, based on statistical losses in training epochs, more parameters may cause more randomness in the process and a minimization of the number of parameters is required to achieve a good reproducibility of the results. To this end, the CNN in the proposed system has a unique structure with 2 distinguished characters. Firstly, the three paths of its feature extraction block are designed to extract, from the multi-modality input, comprehensive feature information of mono-modality, paired-modality and cross-modality data, respectively. Also, it has a particular three-branch classification block to distinguish pixels in each of the 3 intra-tumoral classes from the background. Each branch is trained separately so that the parameters are updated specifically with the corresponding ground truth data of target tumor areas. The convolutional layers of the system are custom-designed with specific purposes, resulting in a very simple config of 61,843 parameters in total. The proposed system has been tested extensively with BraTS2019 and BraTS2018 datasets. The mean Dice scores, obtained from the ten experiments on BraTS2019 validation samples, are 0.751 ± 0.007, 0.885 ± 0.002, 0.776 ± 0.004, for enhancing tumor, whole tumor and tumor core, respectively. The test results demonstrate that the proposed system is able to reproduce a high-quality segmentation result quite consistently. Furthermore, its extremely low computation complexity will facilitate its implementation/application in various environments.

Keywords Brain tumor segmentation, convolutional neural network, multi-path feature extraction and classification block, performance reliability and reproducibility, separate and parallel training.

1. Introduction

Brain tumors cause serious brain diseases and brain tumor detection is important for the diagnosis. In general, brain tumor segmentation is to detect and to localize tumors in 3D brain images. Each tumor area can be further segmented into 3 intra-tumoral structures, namely edema (ED), non-enhancing /necrotic tumor (NET) and enhancing tumor (ET). As the segmentation needs to be voxel-wisely precise, brain tumor

segmentation is, ultimately, to classify the voxels of a brain image into 4 classes, i.e., enhancing tumor (ET), tumor core (TC) comprising ET and NET, whole tumor (WT) comprising TC and ED, and the background. In case of brain scanning by Magnetic Resonance Imaging (MRI), four 3D brain images of 4 modalities, i.e., Flair, T2, T1ce and T1, are acquired for each patient case. Each 3D image is usually sliced into a large number of 2D slices. An example of brain image slices of the 4 modalities and the segmentation result approved by neuroradiologists [1] is illustrated in Fig. 1. The brain tumor segmentation is, in fact, a task of multi-class classification of multi-modality data. It requires (i) examination of a large amount of data, and (ii) good knowledge of neuroradiologists. Hence, manual segmentation is a time-consuming and difficult task usually performed by medical specialists, which may lead to lengthy wait for diagnosis results in many patient cases. Developing fully automated brain tumor segmentation system by computer vision can help to reduce the workload of doctors.

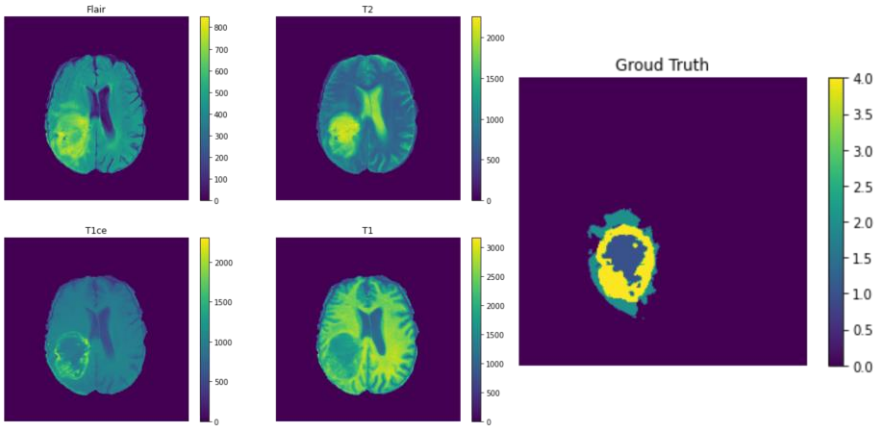

Fig. 1. An example of MRI images of the four modalities, and the ground truth image of its intra-tumoral areas. The ground truth is done manually and approved by medical specialists.

Brain tumor segmentation by computer vision comprises two processes, extraction of the image feature data representing tumor patterns and regrouping the feature data for classification. The feature extraction is done mainly by filtering operations performed sequentially and/or concurrently. Various filters, such as Sobel filters [2] and Gabor filters [3], have been used to extract feature data of different patterns. However, the tumor patterns have a lot of variations, and one cannot apply a large number of filters of handcrafted coefficients to deal with them, which causes a loss of critical information and thus affects the quality of the classification in the succeeding process. The feature data regrouping for classification can also be done by means of filtering operations, but it has also the difficulty in manually selecting a large number of suitable filters to correctly classify the variable tumor areas.

Convolutional neural networks (CNN) have been used effectively for various image processing purposes [4]-[8]. The filtering coefficients in a CNN are determined by means of progressive update in a training process. Thus, one can apply a sufficient number of filters to handle various kinds of feature data, without need to choose them manually. Moreover, as the network is trained with the ground truth data verified by medical specialists, the filters of the network can be set-up and fine-tuned with their knowledge in the topic area.

It should, however, be mentioned that, as a CNN involves a large number of filters, operating the CNN requires a huge number of calculations, which may limit its application in a computation resource restricted environment. Moreover, there is also a problem of reproducibility related to the randomness in the training process.

A training process is used to determine the values of the filtering coefficients in the convolutional layers of a CNN. The coefficient values are updated iteratively, according to the losses generated in the training epochs. There are many sources of randomness in this process. Firstly, each loss is calculated statistically in processing a batch of training samples. The number of samples per batch is usually fixed, but the composition of the batches is different from epoch to epoch in order to have statistical updating results. To this end, the samples in the training pool are randomly shuffled in each epoch. Moreover, if training samples are insufficient, with respect to the number of coefficients to be trained, various data argumentation methods are applied, which can add more randomness in forming the batches and thus more randomness in the loss calculation results. The randomness can also be caused by the initialization of the coefficients, as one uses a particular distribution of the initial values to help the convergence in the updating process. Other elements, such as GPU floating point calculations, can also add more randomness [9].

Due to the randomness, retraining a CNN system under the same training condition will result in a set of filtering coefficients different from that obtained from a previous training. Hence, the system is not the same and cannot reproduce the same results with the same inputs. However, the reproducibility of a system for human health applications is an issue of critical importance [10]. In general, more filtering coefficients being trained in the process can lead to more randomness, and the poorer reproducibility. Therefore, excessively use of filters should be avoided.

In this paper, we propose a dual tri-path, highly reliable and computation efficient CNN system specifically for brain tumor segmentation of MRI brain images. The CNN in the system is custom-designed with the objective that (i) its structure is suitable to process multi-modality input data and to perform multi-class classification, and (ii) its dimensions, i.e., the number of layers and number of filtering kernels per layer, are just-sufficient for the complexity of the task to minimize the randomness in the training while securing the processing quality. To this end, new design approaches are developed to have (a) each filter placed on purpose, (b) the network training optimized, (c) the computation efficiency maximized, and (d) the reproducibility of the system improved. By achieving the objective of the work, it also contributes to the CNN community with the following issues.

- Multi-path feature extraction (FE) block for multi-modality input data. The input data of each modality carry specific feature information and are correlated with the data of the other modalities. The multi-path structure of the block permits each path to be adjusted to optimize the extraction of mono-modality, paired-modality and cross-modality features, respectively.
- Decomposition of a complex multi-classification problem into several binary classification problems by a multi-branch classification block. The config of each branch is trained independently to better identify the patterns of a particular tumor area, and the filtering coefficients can be updated separately and specifically using the corresponding ground truth data.
- High computation-efficiency and reproducibility achieved by custom-designing the CNN system. The proposed system, with 61,843 parameters in total, is capable of producing high-quality segmentation

results in a consistent manner. The standard deviations of the mean Dice scores of WT, TC and ET, obtained from 10 experiments, are less than 1%.

This paper consists of 5 sections. The related work is presented in Section II. The detailed description of the proposed system is found in Section III. Section IV is dedicated to the presentation of the training details and the experiment results. A conclusion is presented in Section V.

2. Related Work

U-net structure [11], composed of a contracting path to extract features and an expanding path with skip connections to perform classification, is widely used in developing CNNs for image segmentation. In particular, to segment 3D medical images, the U-net structure has been adapted with different variations.

As the 3D brain images are often seen as sequences of 2D slices, many segmentation systems involve 2D U-net based CNNs, of which the input and ground truth data are 2D samples. In order to use the inter-slice feature information, there are also 3D CNNs operating with 3D input samples. In the implementation of such 3D CNNs, one of the challenging issues is to get training samples of sufficient quality and quantity. In some cases, the limitation of computation resources, including memory usage in a training process, can also be an issue. If the data samples are 3D images, instead of 2D slices, much more memory space will be needed to keep the data of an entire batch of 3D samples for the calculation of a loss, and the memory usage limitation can be very critical. To bypass such issues, varieties of pseudo-3D CNNs have appeared. Some used the combination of 2D and 3D convolutions where 2D ones were for extracting intra-slice features, while 3D ones were for extracting inter-slice features [12]. Others sliced each 3D input image into three orthogonal axes to obtain 3 sequences, then performed 2D convolutions in order to capture spatial information with low dimensional convolutions [13]-[17].

To enhance the computation in convolutional layers, in many U-net based CNNs, multi-convolutional blocks were used as the basic computational units. For example, a standard convolutional layer was replaced by a residual block [16], [17], inception block [18], [19], or dense block [12], [20]-[22]. In some U-net designs, the skip connections were enhanced by employing attention gates [17], [23], or squeeze-and-excitation blocks [22], [24]. In case of Unet++ [25], simple skip connections were replaced by a network of convolutional units.

The variations of U-net can also be in network structure. One could use multiple contracting paths in order to process the input data of different modalities. To handle the input data of each modality independently, instead of single one in the original U-net, four contracting paths were used to avoid the false-adaptation between the modalities [26]. There's also a U-net with 2 contracting paths to extract features from flair-t2 pair and t1ce-t1 pair respectively in order to capture the discriminative patterns between brain tumor areas and normal brain areas [27].

Since the brain tumor detection is a fine classification of pixels in a hierarchical structure of tumor sub-regions, multiple U-nets are used in such a way that the results of the proceeding U-net are applied to the succeeding one to perform finer segmentation process [28]-[31]. For example, one can use three U-nets to detect WT, TC and ET, successively [29].

It should, however, be noted that the high performance of a CNN system is often achieved at the cost of large computation resources, which limits its application in a computation resource restricted environment

and increases the risk of poor reproducibility. Therefore, the work presented in this paper aims at custom-designing a CNN system that is able to perform high-quality brain tumor segmentation with very low computation cost and moreover to reproduce almost the same results after retraining.

3. Proposed System

The input of the proposed system consists of four 3D MRI images, corresponding to the 4 modalities, namely Flair, T2, T1ce and T1. Each 3D image is sliced into 2D slices. Thus, the voxels in 3D images become pixels in 2D space. The output is a sequence of slices, in which each pixel is labeled to be in one of the 4 classes, i.e., ED, NET, ET and the background. The proposed system consists of three parts:

1. Pre-processing. It is to exclude the excessive margins and evident tumor-free slices, which reduces the regions of non-interest and the data volume to be processed.
2. CNN. It's the main part of the proposed system to perform brain tumor segmentation. It involves a three-path feature extraction block to take fullest advantage of the multi-modality input, a bottleneck to perform tumor localization, and a three-branch classification block to optimize the segmentation.
3. Refinement block. It aims at correcting falsely labeled enhancing tumor pixels to improve segmentation quality.

3.1 Pre-processing and Refinement Blocks

The pre-processing and refinement blocks are placed to facilitate the operations in the CNN and to improve the segmentation results, respectively.

The pre-processing block is to reduce the volume of data to be processed in the CNN, by removing some of the regions that are apparently out of the regions of interest, from the 4 series of slices in each patient case. Two kinds of regions of non-interest are targeted.

- Excessive margins in each slice. In case of samples from BraTS datasets, one can crop each slice to reduce its dimensions from 240×240 to 200×168 pixels, cutting off more than 40% of the input data in terms of pixels for the CNN, without losing any pixels in the brain areas.
- Tumor-free slices located in the two ends of each sequence. In case of BraTS samples, the first fifteen slices and the last twelve slices do not contain any tumor areas and are thus removed.

The removal of the 2 kinds of regions of non-interest by the pre-processing block results in a significant data reduction. In terms of BraTS samples, it can reach 52% in terms of number of pixels, which helps not only to reduce the computation burden in the succeeding CNN, but also to increase the information density of its input data.

The refinement block is to verify if any of the pixels placed, by the CNN, in the ET areas are falsely classified. As an enhancing tumor area is a 3D object, the pixels of the area are expected to appear in a certain number of consecutive slices. The distribution of enhancing tumor thickness, illustrated in Fig. 2, shows that such a tumor area appears in at least 6 consecutive slices, corresponding approximately to 6 mm in length [1]. Also, as enhancing tumor areas are unlikely to be thin-cylinder shaped, the total number of pixels in such an area should be above a certain threshold. In this work, this threshold is 1000 [32] to have a reasonable

average size of cross-section areas in the 6 or more slices.

The refinement operation is effective to detect false positive enhancing tumor areas. In our experiments, with the 2 thresholds, i.e., 6 consecutive slices and 1000 pixels, the false positive rate, defined as the ratio of the number of false positive voxels to the total number of predicted positive voxels, is reduced from 36% to 29%, without increasing the false negative rate.

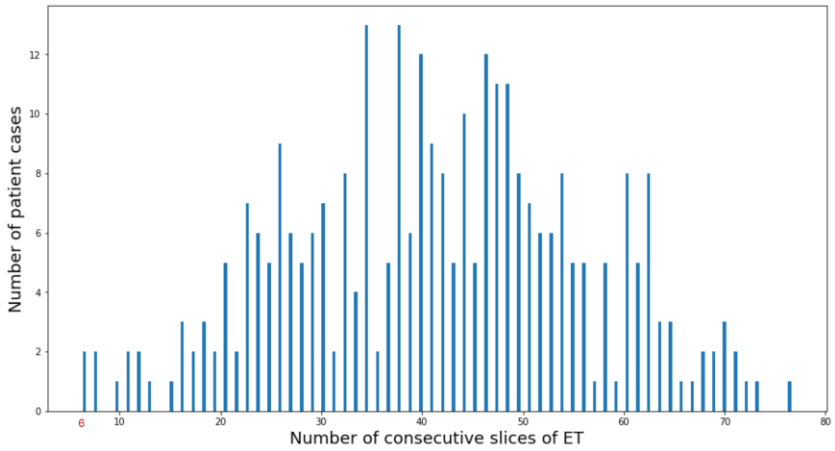

Fig. 2. Distribution of enhancing tumor thickness in number of consecutive slices. The 335 data samples are from the training pool of BraTS 2019 dataset.

3.2 CNN

The CNN proposed in this paper is the major part of the system for the brain tumor segmentation. The emphasis in the design of the network is on the processing quality, reproducibility after retraining, and computation efficiency, i.e., maximizing the processing quality while minimizing the computation cost and randomness.

It should be noticed that segmenting a brain image, in a pixel-wise precision, into tumor areas and tumor-free areas is, in fact, a binary classification operation. As each pixel in a tumor area needs to be further identified which intra-tumoral areas it belongs to, the brain tumor segmentation is a pixel classification of multiple classes. To do so, one needs first to get high-level feature data representing the sematic information in a large neighborhood surrounding each pixel, and then to process the data to see how much likely the pixel is located in any of the tumoral areas. Hence the first part of the CNN is used for feature extraction and the last one for pixel classification.

The detailed architecture of the proposed dual tri-path CNN is shown in Fig. 3. It consists of the following three blocks.

1. Feature extraction (FE) block. It has three-path convolutions, which are dedicated to extracting different features from the four-modality input.
2. Bottleneck. It is to generate feature maps representing tumor locations of low resolution.
3. Three-branch classification block. Each of the branches is made to identify the pixels of a specific kind of tumor areas.

The design details of the three blocks are presented in the following subsections.

3.2.1 Feature Extraction Block

The feature extraction (FE) block is designed to extract comprehensive image features from 2D MRI image slices of the 4 modalities, namely Flair, T2, T1ce and T1. A set of 4 slices from the 4 modalities is illustrated in the first row of Fig. 4. The brain images are acquired by means of the 4 modalities to enhance the appearances of various kinds of pathological tissues to facilitate diagnosis. For example, the features of WT are more identifiable in Flair and T2, whereas the T1ce and T1 scans are tuned to enhance the appearances of ET and TC [1]. As the intensity ranges in the scans of different modalities are not uniformed, the input data need an appropriate normalization in order to facilitate the following filtering operations.

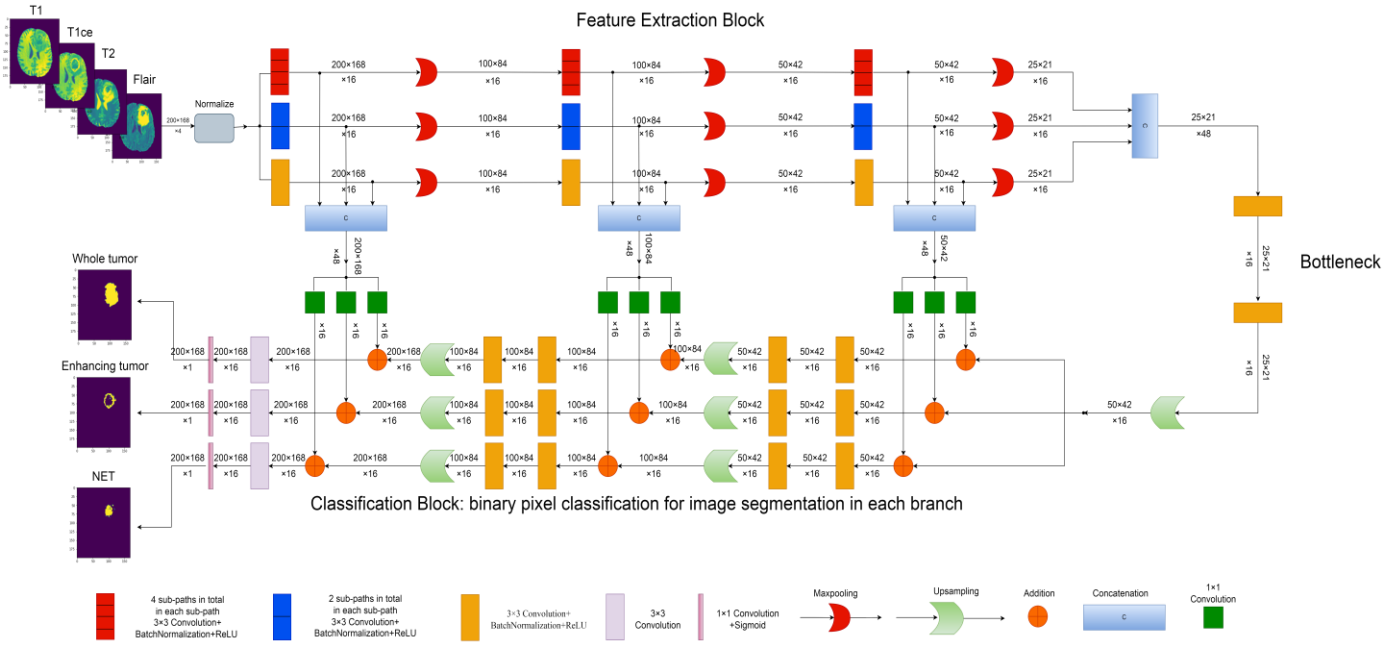

Fig. 3. Detailed block diagram of the proposed CNN. The 3-path Feature Extraction (FE) Block is for the comprehensive feature extraction. The skip connections, going vertically from Feature Extraction Block to Classification Block, are designed to ensure that the features extracted in each of the 3 FE paths are received by all the 3 branches of Classification Block. In each of these 3 branches, a binary classification is performed to identify pixels in one of the 3 intra-tumoral areas. The outputs of the 3 branches are then combined to result in a segmentation of a brain image in a pixel-wise precision.

The aforementioned normalization will be applied to the 3D data of each modality. The mean and standard deviation is calculated from the pixels in the brain areas of all input slices, and those in the background with no region of interest are excluded. As a result, the mean and standard deviation obtained can better represent the distribution of data within the brain area, and consequently, certain features in the brain areas can be enhanced and the contrast of image can be improved as shown in Fig. 4.

The purpose of the comprehensive feature extraction is to capture the data representing critical pathological patterns from brain images. Considering the different emphasis on the varieties of tissues, the input data of each of the modalities need to be filtered individually in order to make good use of their feature characters. However, as the data of all the 4 modalities are acquired from the same source and therefore correlated, they also need to be processed collectively.

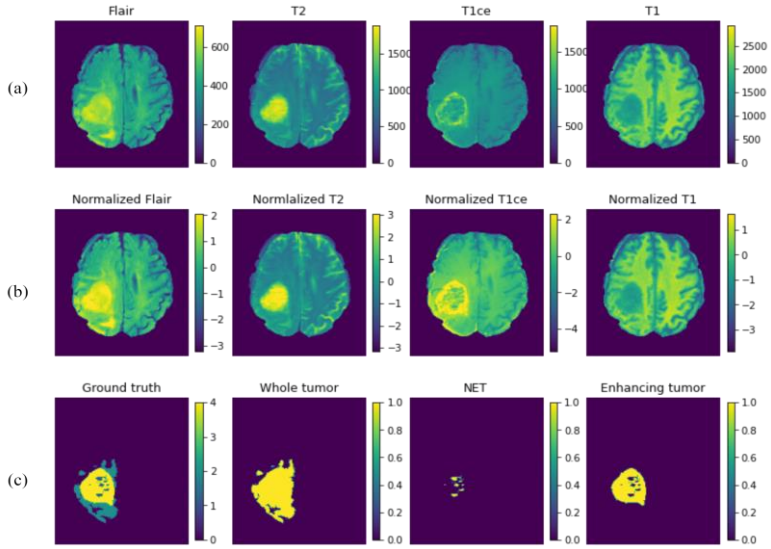

Fig. 4. (a) Input slices of the 4 modalities, (b) normalization results, and (c) ground truth images.

It is known, on one hand, that the patterns of the whole tumor region are more distinguishable in Flair and T2, and one can get more whole tumor features by means of convolutions applied to these 2 modalities combined. On the other hand, the pair of T1 and T1ce display more differences between the patterns in tumor core and those in the rest of brain areas, and filtering them together can produce high-quality feature data representing tumor cores and enhancing tumors.

Taking the above-mentioned points into consideration, the FE block is designed with 3 paths, as shown in Fig. 3. The input data of all the 4 modalities are applied to all of the 3 paths, and get convolved 3 times in each path. The input data are organized, however, into 3 different groupings so that the 3 paths can produce the following 3 kinds of features.

- Mono-modality features. The upper path of the FE block has 4 sub-paths, in each of which the input data of mono-modality are processed in the 3 successive layers of convolution & max-pooling operations. Each of the convolutions has 4 kernels, producing adequately 4 data maps per modality. The 4 sub-paths generate, in total, 16 mono-modality feature maps from the input data of Flair, T2, T1 and T1ce.
- Paired-modality features. In the mid path, there are 2 sub-paths, where the input data of Flair-T2 are applied to one sub-path for whole tumor features, and those of T1-T1ce are applied to the other for tumor core and enhancing tumor features. Each convolution has 8 kernels, considering that it involves the data of 2 modalities. There are 16 feature maps generated by the 2 sub-paths.
- Cross-modality features. There is no separated sub-path in the lower path of the FE block. The 3 successive convolution & max-pooling operations are applied to the data of all the 4 modalities combined. This path produces 16 cross-modality feature maps.

A max-pooling operation following each convolution is performed to enlarge the receptive field and to increase the information density, which, nevertheless, reduces the resolution of the feature maps. The 3-path feature extraction block generates, in total, 48 high-sematic and low-resolution feature maps. They carry comprehensive feature information of cross-modality, individual modality, and paired modality for further processing.

3.2.2 Bottleneck

As shown in Fig. 3, the bottleneck is located between the 3-path FE block and the 3-branch classification block. Its function is to transform the input data, i.e., 48 low-resolution feature maps, into data maps indicating the potential tumor locations in order to perform coarse segmentation.

The function of the block is implemented by means of two simple convolutional layers. The number of kernels per layer is modestly 16 to minimize the risk of introducing randomness, with a view to achieving a good reproducibility of the system. Since the input data maps are of low resolution and convolutions do not increase the resolution. The output maps can carry the information of potential tumor locations, without details of tumor areas. The information will be used in the classification block, presented in the following subsection.

3.2.3 Classification Block

The classification block, as shown in Fig. 3, is to classify all the pixels into 4 classes, i.e., WT (comprising ED, ET, NET), ET, NET, and the background. The ET and NET pixels form the TC areas. The prime input of this block is a set of low-resolution data maps produced by the bottleneck.

The basic processing elements in this block are convolution, upsampling and skip connection. The purpose of the upsampling, by means of bilinear interpolation, is to gradually restore the initial image dimension. Because the fine classification needs detailed image information which is lost due to the three max-pooling operations in the feature extraction block, the feature maps produced before each max-pooling operation are brought in by means of the skip connections. They are scaled with learnable coefficients, and then used to modulate the upsampled data maps. Two layers of convolutions are employed to fine-tune the modulated signal in order to get precise segmentation result.

As mentioned earlier, there are 3 kinds of tumor areas needed to be identified, i.e., WT, ET and NET. They are different in sizes, textural patterns, and intensity ranges. The areas of whole tumor, enhancing tumor and NET in a slice of a patient case are illustrated in Fig. 5 as examples. By examining images of patient cases, one can have the following observations.

- Whole tumor areas include various pathological tissues. The have relatively large dimensions. Although the texture patterns have a lot of variations inside the areas, they are different from those outside the areas. Identifying whole areas from a brain image is essentially to distinguish the patterns of pathological tissues from those of the normal brain structures.
- The other kinds of tumor areas, namely ET, NET and edema, have patterns different from one another. If a set of filters is made to detect one kind of tumor areas, it may not be suitable for the others.

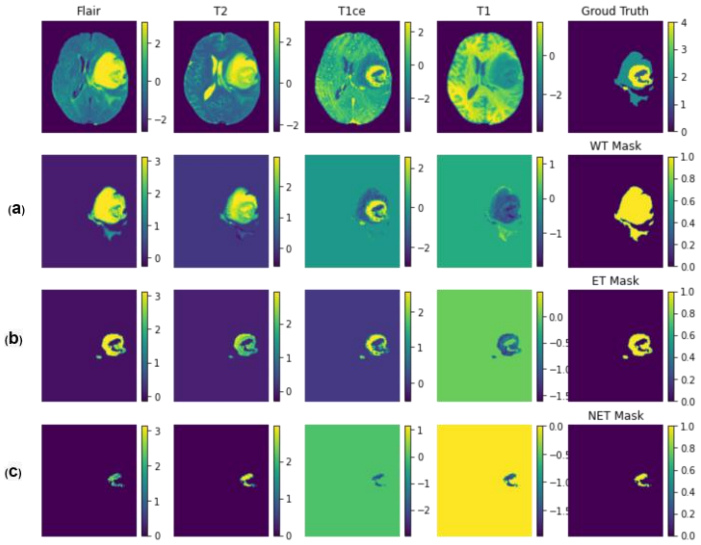

Fig. 5. Example of WT, ET and NET areas and their masks. (a) WT area in Flair, T2, T1ce and T1, respectively. (b) ET area in Flair, T2, T1ce and T1, respectively. (c) NET area in Flair, T2, T1ce and T1, respectively.

In light of above observation, to precisely classify the pixels in the 3 different kinds of tumor areas, different criteria to differentiate image patterns are needed, which may not be easily done by a single filtering path. Hence, the classification block of the proposed system is designed to have 3 branches in parallel, as shown in Fig. 3, to identify the pixels in the whole tumor (WT), enhancing tumor (ET), and NET areas, respectively. In other words, each branch is specifically made to perform a binary classification, differentiating the pixels in one of the 3 kinds of tumor areas and the rest of the brain image. For example, the second branch, shown in Fig. 3, is dedicated to the binary classification of ET pixels versus non-ET pixel, i.e., the rest of the pixel population of the brain image. As the locations of the ET pixels are found in this binary classification, the ET areas are precisely determined in the brain image and an ET mask is then generated. The operations in the WT and NET branches are expected to result in NET and WT masks, respectively. As the tumor core (TC) areas comprise ET and NET areas, the TC mask can be obtained by combining the ET and NET masks. Therefore, a complex classification of 3 intra-tumoral classes versus the background is decomposed into 3 simple binary classifications. The brain tumor segmentation can thus be done by means of the 3 binary pixelwise classifications performed, respectively, in the 3 separated branches of the classification block.

The three branches are configured identically, as shown in Fig. 3, and each has three sets of convolutions. In each set, the upsampled data maps are modulated by the feature maps of the same dimensions from the feature extraction block. To optimize the modulation result, the feature maps are appropriately scaled by means of 1x1 convolutions. The modulated data maps are then convolved to perform a precise detection of targeted tumor area.

In order to make each of the 3 branches capable of identifying one specific kind of tumor areas, they are trained independently with the binary masks of the whole tumor, enhancing tumor and NET, respectively. As a result, the filtering coefficients can be adjusted to suit its own pattern detection.

The final output of the classification block is three feature maps indicating the areas of whole tumor, enhancing tumor and NET in the original input.

3.2.4 Network Configuration

The detailed network configuration is shown in Table 1. The total number of parameters is only 61,843, which is merely a very small percentage of that of many CNNs found in published research papers.

Table 1 Details of the network configuration

Layer	Kernel Size	Input size	Output Size	Number of kernels	Number of Parameters
1	3×3	200×168	100×84	48	1,248
2	3×3	100×84	50×42	48	4,272
3	3×3	50×42	25×21	48	4,272
4	3×3	25×21	25×21	16	6,992
5	3×3	25×21	25×21	16	2,384
6	3×3	25×21	50×42	48	7,152
7	3×3	50×42	50×42	48	7,152
8	3×3	50×42	100×84	48	7,152
9	3×3	100×84	100×84	48	7,152
10	3×3	100×84	200×168	48	6,960
11	1×1	200×168	200×168	3	51
SC1*	1×1	25×21	25×21	48	2,320
SC2*	1×1	50×42	50×42	48	2,320
SC3*	1×1	100×84	100×84	48	2,320
Total					**61.843**

*SC: skip connection

In general, a dual tri-path, computation-efficient CNN system, combined with pre-processing and refinement block, is custom-designed to perform high-quality brain tumor segmentation based on the analysis of the characters of the input data and the output requirements. The proposed network has two highlights:

1. The feature extraction block has 3 paths and each of them is to extract comprehensive features of mono-modality, paired-modality and cross-modality respectively from the multi-modality input data.
2. There are three branches in the classification block, which decomposes a complex multi-class classification problem into three simple binary classification tasks. This structure enables an appropriate training of these branches so that each of them can be adjusted effectively to suit the classification of the pixels of the specific tumor areas.

4. Performance Evaluation

The performance evaluation of the proposed system has been conducted on the datasets of BraTS 2018 and BraTS 2019. The assessment of the validation results has been done by means of CBICA Image Processing Portal [33], [34]. Two bunches of experiments have been carried out to assess the quality of the proposed system in different aspects. The first bunch, consisting of 10 experiments, has been to evaluate the reproducibility and performance consistency, which is presented in Subsection 4.4. The second bunch has been done for the ablation study, and is described in Subsection 4.5. Subsection 4.6 is dedicated to the

presentation of the performance comparison between the proposed system and those recently reported in reputed research journals.

4.1 Dataset

BraTS 2019 is used to train the system and also for the performance evaluation. There are 335 patient cases in the training pool and additional 125 cases for validation. These data samples are multi-institutional routine clinically-acquired pre-operative multimodal MRI scans. All the cases are of pathologically confirmed diagnosis with available overall survival (OS). All the images have been segmented manually, by one to four raters, following the same annotation protocol, and their annotations were approved by experienced neuroradiologists. All the ground truth labels have been manually-revised by expert board-certified neuroradiologists [35].

A simple data augmentation, by flipping and rotating, is applied to the training samples to increase their number for a better training. Moreover, the training samples are shuffled in each epoch so that they are rearranged differently in batches.

4.2 Implementation Details

The hyperparameters determine, to some extent, how the filtering coefficients of the proposed CNN are derived. For the training process, the most important hyperparameters are as follows.

1. Batch size. There are 335 patient cases for training, i.e., 335×155 slices per modality. Taking this number of slices and the limit of memory usage into consideration, each batch is made to have 128 slices.
2. Number of epochs. The training requires only 75 epochs for the loss value to reach its minimum level.
3. Learning rate. It is variable with staircase decay. The initial learning rate is 0.005 and will be reduced by half every 15 epochs.
4. Loss function. Dice loss function is applied to calculate the loss in the WT branch. For NET and ET branches, Tversky loss [36] is chosen to reduce the risk of divergence caused by small-size targets.
5. Optimizer. Adam [37] is chosen as the optimizer for the training process.
6. Kernel initialization. The kernel coefficients are initialized with a uniform distribution by he_uniform [38].

Tensorflow is used to implement the proposed system. The total training time is 7 hours for BraTS 2018 dataset and 9 hours for BraTS 2019 dataset with NVIDIA Tesla P100 GPU.

4.3 Evaluation Metrics

The segmentation result is mainly evaluated by 4 metrics: Dice score, Sensitivity, Specificity and Hausdorff95. Let P_0 and P_1denote the predicted results of tumor-free and tumor regions, T_0 and T_1 are the ground truth of tumor-free and tumor regions. Then, the first three metrics are defined as:

$$Dice(P_1, T_1) = \frac{P_1 \cap T_1}{(P_1 + T_1)/2} \quad (1)$$

$$Sensitivity(P_1, T_1) = \frac{P_1 \cap T_1}{T_1} \quad (2)$$

$$Specificity(P_0, T_0) = \frac{P_0 \cap T_0}{T_0} \quad (3)$$

Hausdorff95 is 95^{th}percentile of the maximum distance of a point on the predicted result to the nearest point on the ground truth. Dice score is the most important and comprehensive metric for evaluating brain tumor segmentation result. It measures volumetric overlap between segmentation results and ground truth.

4.4 Reproducibility Study

The proposed system is developed for a medical application, a good reproducibility is required for it to be usable in practice. As mentioned previously, the filtering coefficients of a CNN system are determined in a training process, in which there is some randomness in calculating the training loss and parameter updating. This randomness causes a reproducibility problem, i.e., the system not being able to reproduce the same result after being retrained under the same training conditions.

A good system should have a good, though not perfect, reproducibility. To measure the reproducibility of the proposed CNN system, we have conducted 10 experiments on the BraTS 2019 dataset. In each of them, the 61,843 parameters of the system are reinitialized, and then retrained with the data from the same training pool. After that, the retrained system is tested with the 125 samples from BraTS 2019 validation pool. Hence, each experiment results in 125 sets of ET, WT and TC Dice scores. The mean and median Dice scores obtained in each experiment, are presented in Table 2. One can have the following observations concerning the reproducibility of the proposed system.

- The system performs a high-quality segmentation. For example, the mean Dice scores of ET in the 10 experiments are ranged from 0.745 to 0.767. The mean Dice scores of TC are ranged from 0.774 to 0.784. The mean Dice scores of WT are ranged from 0.883 to 0.886.
- The average mean Dice scores of WT, TC and ET, obtained from the 10 experiments, have a high degree of consistency. The standard deviations are all below 1%.
- The median Dice scores of the 10 experiments, found in Table 2, also demonstrate a similar degree of performance excellence and a good reproducibility of the proposed system.

It is also important to measure the reproducibility of the proposed system in processing individual patient cases. A reliable system should be able to reproduce similar results for a given patient case after each training run (i.e., experiment). The Dice scores obtained from 10 experiments of 2 typical cases are presented in Table 3, demonstrating that all the 10 experiments result in the 10 systems are very similar to each other and thus they reproduce the test results with very small deviations.

The experiment results demonstrate that the proposed system is able to operate in a consistent manner and reproduce fine segmentation results after each retraining, which makes it useful for medical applications. The results also give a confirmation that the measures taken, in the design of the proposed CNN, to reduce the

network randomness are effective. It also proves that a high degree of processing quality and reproducibility can be achieved all together by an approach of custom-designing CNN systems for target-specific applications.

Table 2 Results of the 10 experiments on BraTS 2019 validation set

Exp	Dice-Mean			Dice-Median		
	ET	WT	TC	ET	WT	TC
1	0.748	0.884	0.776	0.843	0.912	0.860
2	0.767	0.886	0.774	0.861	0.913	0.859
3	0.747	0.883	0.771	0.848	0.912	0.861
4	0.747	0.882	0.776	0.842	0.911	0.850
5	0.754	0.889	0.779	0.844	0.916	0.866
6	0.753	0.885	0.775	0.849	0.910	0.846
7	0.751	0.884	0.768	0.856	0.916	0.857
8	0.752	0.882	0.776	0.844	0.915	0.869
9	0.745	0.883	0.784	0.854	0.917	0.862
10	0.746	0.887	0.779	0.841	0.915	0.862
Best case	**0.767**	**0.886**	**0.774**	**0.861**	**0.913**	**0.859**
Worst case	**0.745**	**0.883**	**0.784**	**0.854**	**0.917**	**0.862**
Average	**0.751**	**0.885**	**0.776**	**0.848**	**0.914**	**0.859**
STDDEV	**0.007**	**0.002**	**0.004**	**0.006**	**0.002**	**0.007**

Table 3 Dice scores of two patient cases

Exp	Case 1*			Case2**		
	ET	WT	TC	ET	WT	TC
1	0.747	0.715	0.645	0.859	0.857	0.711
2	0.765	0.708	0.649	0.854	0.871	0.767
3	0.780	0.715	0.636	0.861	0.851	0.725
4	0.738	0.686	0.498	0.855	0.862	0.831
5	0.770	0.712	0.543	0.850	0.861	0.850
6	0.776	0.696	0.487	0.867	0.867	0.790
7	0.756	0.718	0.586	0.850	0.860	0.732
8	0.672	0.754	0.684	0.861	0.869	0.857
9	0.747	0.713	0.614	0.857	0.858	0.737
10	0.783	0.703	0.605	0.862	0.874	0.880
Average	**0.753**	**0.712**	**0.595**	**0.858**	**0.863**	**0.788**
STDDEV	**0.031**	**0.017**	**0.063**	**0.005**	**0.007**	**0.059**

* Case 1: BraTS19_WashU_W047_1

** Case 2: BraTS19_TCIA13_611_1

4.5 Ablation Study

The proposed system is designed specifically for brain tumor segmentation and has a number of specific characters, in network configuration and in model parameters, which differentiates it from the other CNN systems for the same task. An ablation study is carried out to assess their effectiveness in processing quality.

The CNN in the proposed system has 4 characters:

1. three-path extraction of mono-modality, paired-modality and cross-modality feature data,

2. classification of 4-class pixels by 3 convolutional branches allowing separated training of the filtering coefficients,
3. identification of WT, NET and ET pixels, instead of WT, TC and ET, and
4. minimized but sufficient numbers of kernels in convolutional layers.

Four tests are performed on BraTS 2019 dataset to measure the effectiveness of each of the 4 characters in achieving a good segmentation quality. The results are presented in Table 4, and each value presented is a mean score obtained in 3 repeated experiments.

Table 4 Ablation study results on BraTS 2019 validation samples

Tests	Ablation type	Dice			Sensitivity			Number of parameters
		ET	WT	TC	ET	WT	TC	
Test1	Single-path FE block	0.708	0.881	0.771	0.731	0.877	0.779	95.971K
Test2	Single-branch classification	0.732	0.882	0.773	0.762	0.889	0.784	115.667K
Test3	ET/ WT/TC classification	0.742	0.885	0.777	0.787	0.888	0.772	61.843K
Test4.1	Doubled number of convolution kernels	0.750	0.883	0.777	0.792	0.892	0.773	241.187K
Test4.2	Halved number of convolution kernels	0.721	0.874	0.753	0.765	0.878	0.760	16.235K
Proposed CNN		**0.751**	**0.885**	**0.776**	**0.796**	**0.891**	**0.783**	**61.843K**

In the first test, the 3-path FE block is replaced by a single path as shown in Fig. 6, involving 3 layers of 48 kernels per layer, and the rest of the network remains unchanged. This replacement, while increasing the total number of parameters from 61.843K to 95.971K, results in a decrease of Dice scores in all the tumor categories. In particular, the decrease of ET Dice scores is even more significant, i.e., from 0.751 to 0.708. It has been confirmed that the 3-path convolution block effectively optimizes the feature extraction from the 4-modality brain images, and it is thus able to produce high-density features information that leads to a better segmentation.

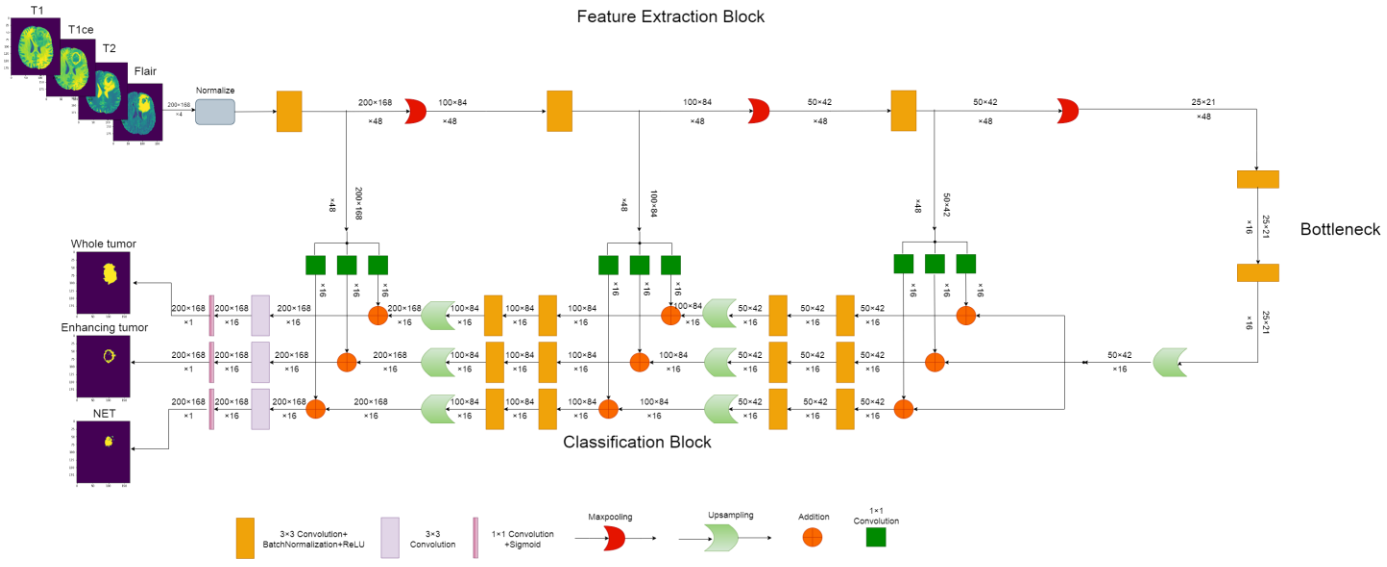

Fig. 6. Block diagram of the CNN used for Test 1 of Ablation Study. The Feature Extraction Block has only one path that generates 48 feature maps, whereas in the proposed CNN, shown in Fig. 3, there are 3 separated FE paths and each generates 16 feature maps.

The second test is carried out to measure the effectiveness of the 3 branches in the classification block. For this measurement, the system is modified in such a way that the 3 branches are combined into single one as shown in Fig. 7. The number of kernels in each layer is tripled to be 48, which increases considerably the number of parameters from 61.843K to 115.667K. This modification makes the segmentation less accurate, as shown by the data presented in the second row in Table 4, and the most noticeable is the decrease of 2% in ET Dice score. This test proves that, in the proposed system, the 3-branch structure for the classification of the pixels in 3 different tumor categories is effectively advantageous in terms of processing quality and computation cost. While classification of multi-class is considered as a complex task, this structure permits each branch to be trained separately and specifically to handle a binary classification of a particular tumor category, which simplifies the task and improves the processing quality.

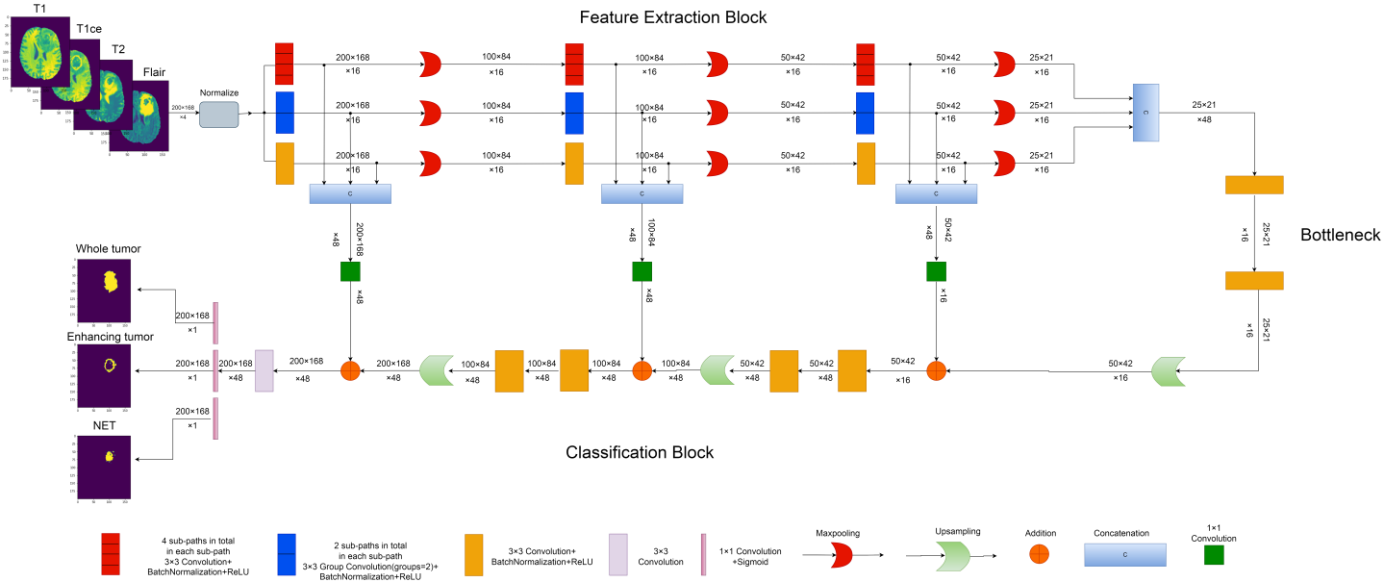

Fig. 7. Block diagram of the CNN used for Test 2 of Ablation Study. It is obtained by merging the three branches in the classification block of the proposed CNN into one to perform a multi-class pixel classification, instead of the 3 separated binary classifications.

In the third test, the three classification branches are dedicated to identifying pixels of WT, TC and ET, instead of WT, NET and ET in the proposed system. As shown in Table 4, the Dice score and sensitivity of ET are both decreased by about 1%, indicating that identifying pixels in NET areas helps the segmentation of ET and TC areas.

The fourth test is to illustrate that the proposed system, despite its very small number of parameters, is able to perform high-quality segmentation of brain images and the number of parameters is just-sufficient. The test has been conducted in the following 2 phases.

- In Phase 1, the number of kernels of each convolution is doubled from 16 to 32, making the total number of parameters increase from 61.843 K to 241.187K. The test results are presented in the 4th row in Table 4. They have demonstrated that increasing the number of convolution kernels did not result in a better segmentation.
- In phase 2, the number of kernels of each convolution is decreased from 16 to 8, reducing significantly the number of parameters, which causes a decrease in processing quality.

The results of these 2 test phases demonstrate that the number of parameters of the proposed system is just-

sufficient. Decreasing it would lead to a degradation of the processing quality, whereas increasing it would require more computation burden without any gain in processing quality.

4.6 Performance Comparison

The performance of the proposed system is compared, in terms of segmentation quality and computation complexity, with that of those recently reported in reputed journals. As some of these results were obtained with BraTS 2019 dataset but others with BraTS 2018 dataset, the performance comparison has been done in the 2 settings: The results of the training/test on BraTS 2019 dataset are presented in Table 5 and those of training/test on BraTS 2018 dataset are found in Table 6.

In the two tables, the results of the proposed system are presented in the bottom rows. The Dice and Sensitivity scores are generated by the 10 experiments specified in Subsection 4.4. One can have the following two observations.

In terms of segmentation quality, the performance of the proposed system is among the best. Its Dice and Sensitivity scores are higher those that of many other CNN systems reported in research journals, with validation samples of both BraTS 2019 and 2018 datasets.

In terms of computation cost, the proposed system is the best, standing ahead of the others by a considerable distance. The entire convolutional network has only 61,843 parameters, while most existing systems need multi-millions. If the number of voxels of each patient case is 240×240×155×4, the number of Flops to complete the test is merely 146G.

The test results demonstrate that the proposed system is able to perform brain tumor segmentation of excellent quality in a consistent manner, and this performance is achieved at a computation cost that is only a small fraction of the average amount of the others. It confirms the effectiveness of the specific characters in the design. They altogether make the proposed system operate with a high computation efficiency, low randomness, and reliable performance.

Table 5 Comparison of the results on BraTS 2019 validation samples

System	Dice			Sensitivity			Number of
	ET	WT	TC	ET	WT	TC	parameters
Di Ieva, *et al.* 2021 [44]	0.675	0.870	0.711	0.702	0.823	0.670	28.03M*
Rehman *et al.* 2021 [42]	0.708	0.869	0.775	N.A.	N.A.	N.A	31.4M*
Zhang *et al.* 2020 [23]	0.709	0.870	0.777	N.A.	N.A.	N.A	7.5M*
Huang *et al.*2021 [39]	0.730	0.827	0.788	0.798	0.967	0.801	N.A.
Ali *et al.* 2021 [45]	0.740	0.880	0.810	0.770	0.890	0.790	N.A.
Liu *et al.* 2021 [40]	0.759	0.885	0.851	N.A.	N.A.	N.A	3.34M
Sun *et al.* 2020 [41]	0.761	0.890	0.779	0.767	0.883	0.762	27.42M*
Best of the 10 mean scores# of the proposed system	**0.767**	**0.886**	**0.774**	**0.800**	**0.889**	**0.784**	**61.843K**
Average of the 10 mean scores# of the proposed system	**0.751**	**0.885**	**0.776**	**0.796**	**0.891**	**0.783**	**61.834K**
Standard deviation	**0.007**	**0.002**	**0.004**	**0.006**	**0.002**	**0.007**	**61.834K**

*Number estimated based on the detailed network configs found in the paper.

Each mean score has been obtained in an experiment with 125 validation cases. The details are found in Subsection 4.4.

Table 6 Comparison of the results on BraTS 2018 validation samples

Systems	Dice			Sensitivity			Number of parameters
	ET	WT	TC	ET	WT	TC	
Huang *et al.* 2021 [39]	0.717	0.801	0.759	0.829	0.962	0.800	N.A.
Akil *et al.* 2020 [20]	0.732	0.860	0.733	0.740	0.838	0.702	181k
Liu *et al.* 2021 [40]	0.767	0.898	0.834	N.A.	N.A.	N.A.	3.34M
Sun *et al.* 2021 [41]	0.771	0.900	0.795	0.769	0.904	0.751	27.42M*
Zhang *et al.* 2020 [23]	0.772	0.872	0.808	N.A.	N.A.	N.A	7.5M*
Rehman *et al.* 2021 [42]	0.773	0.894	0.826	N.A.	N.A.	N.A	31.4M*
Zhang *et al.* 2020[43]	0.782	0.896	0.824	N.A.	N.A.	N.A	363k
Best of the 10 mean scores # of the proposed system	**0.792**	**0.887**	**0.813**	**0.851**	**0.901**	**0.820**	**61.834k**
Average of the 10 mean scores # of the proposed system	**0.787**	**0.886**	**0.801**	**0.837**	**0.900**	**0.816**	**61.834k**
Standard deviation	**0.003**	**0.002**	**0.007**	**0.007**	**0.003**	**0.005**	**61.834k**

*Number estimated based on the detailed network configs found in the paper

Each mean score is obtained in an experiment with 66 validation cases.

5. Conclusion

In this paper, a dual tri-path, computation-efficient and performance consistent CNN system is proposed for brain tumor segmentation. Its CNN is custom-designed on the basis of the analysis of input signal and output requirements. It consists of three parts, (i) a three-path feature extraction block to extract comprehensive features from the multi-modality input, (ii) a bottleneck to perform tumor localization and coarse segmentation, and (iii) a classification block with three branches to detect different tumor areas separately, where each branch is trained independently so that the parameters in each branch can be updated accordingly to suit the detection of specific tumor patterns in order to improve the segmentation accuracy.

The performance of the proposed system has been evaluated on both BraTS 2018 and BraTS 2019 datasets. The results of a good number of experiments demonstrate that the proposed system is able to perform a high-quality segmentation and to reproduce almost the same results in a consistent manner. The comparison with those recently reported in reputed research journals shows that the segmentation quality of the proposed system is among the best while the computation cost is the lowest, merely a small percentage of those of other systems.

Acknowledgement

This work was supported in part by Compute Canada and in part by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

[1] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., ... & Van Leemput, K. (2014). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE transactions on medical imaging*, *34*(10), 1993-2024. P. Gibbs, D. L. Buckley, S. J. Blackband, and A. Horsman, "Tumour volume determination from MR images by morphological segmentation," Physics in Medicine & Biology, vol. 41, no. 11, p. 2437, Nov. 1996.

[2] Gibbs, P., Buckley, D. L., Blackband, S. J., & Horsman, A. (1996). Tumour volume determination from MR images by morphological segmentation. *Physics in Medicine & Biology*, *41*(11), 2437.

[3] Praveen, G. B., & Agrawal, A. (2015, November). Hybrid approach for brain tumor detection and classification in magnetic resonance images. In *2015 Communication, Control and Intelligent Systems (CCIS)* (pp. 162-166). IEEE.

[4] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

[5] Matsugu, M., Mori, K., Mitari, Y., & Kaneda, Y. (2003). Subject independent facial expression recognition with robust face detection using a convolutional neural network. *Neural Networks*, *16*(5-6), 555-559.

[6] Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

[7] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[8] Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

[9] https://www.determined.ai/blog/reproducibility-in-ml

[10] McDermott, M. B., Wang, S., Marinsek, N., Ranganath, R., Foschini, L., & Ghassemi, M. (2021). Reproducibility in machine learning for health research: Still a ways to go. *Science Translational Medicine*, *13*(586), eabb1655.

[11] Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention* (pp. 234-241). Springer, Cham.

[12] Li, X., Chen, H., Qi, X., Dou, Q., Fu, C. W., & Heng, P. A. (2018). H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE transactions on medical imaging*, *37*(12), 2663-2674.

[13] Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Auto-context convolutional neural network (auto-net) for brain extraction in magnetic resonance imaging. *IEEE transactions on medical imaging*, *36*(11), 2319-2330.

[14] Rosas-Gonzalez, S., Birgui-Sekou, T., Hidane, M., Zemmoura, I., & Tauber, C. (2021). Asymmetric ensemble of asymmetric U-Net models for brain tumor segmentation with uncertainty estimation. *Frontiers in Neurology*, *12*, 609646.

[15] Luo, Z., Jia, Z., Yuan, Z., & Peng, J. (2020). HDC-Net: Hierarchical decoupled convolution network for brain tumor segmentation. *IEEE Journal of Biomedical and Health Informatics*, *25*(3), 737-745.

[16] Maji, D., Sigedar, P., & Singh, M. (2022). Attention Res-UNet with guided decoder for semantic segmentation of brain tumors. *Biomedical Signal Processing and Control*, *71*, 103077.

[17] Zhou, T., Ruan, S., Vera, P., & Canu, S. (2022). A Tri-Attention fusion guided multi-modal segmentation network. *Pattern Recognition*, *124*, 108417.

[18] Chen, Y., Cao, Z., Cao, C., Yang, J., & Zhang, J. (2018, June). A modified U-Net for brain Mr image segmentation. In *International Conference on Cloud Computing and Security* (pp. 233-242). Springer, Cham.

[19] Qamar, S., Ahmad, P., & Shen, L. (2020, October). Hi-net: Hyperdense inception 3d unet for brain tumor segmentation. In *International MICCAI Brainlesion Workshop* (pp. 50-57). Springer, Cham.

[20] Akil, M., Saouli, R., & Kachouri, R. (2020). Fully automatic brain tumor segmentation with deep learning-based selective attention using overlapping patches and multi-class weighted cross-entropy. *Medical image analysis*, *63*, 101692.

[21] Yogananda, C. G. B., Shah, B. R., Vejdani-Jahromi, M., Nalawade, S. S., Murugesan, G. K., Yu, F. F., ... & Maldjian, J. A. (2020). A fully automated deep learning network for brain tumor segmentation. *Tomography*, *6*(2), 186-193.

[22] Wang, L., Wang, S., Chen, R., Qu, X., Chen, Y., Huang, S., & Liu, C. (2019). Nested dilation networks for brain tumor segmentation based on magnetic resonance imaging. *Frontiers in Neuroscience*, *13*, 285.

[23] Zhang, J., Jiang, Z., Dong, J., Hou, Y., & Liu, B. (2020). Attention gate resU-Net for automatic MRI brain tumor segmentation. *IEEE Access*, *8*, 58533-58545.

[24] Huang, Z., Zhao, Y., Liu, Y., & Song, G. (2021). GCAUNet: A group cross-channel attention residual UNet for slice based brain tumor segmentation. *Biomedical Signal Processing and Control*, *70*, 102958.

[25] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., & Liang, J. (2019). Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, *39*(6), 1856-1867.

[26] Zhou, T., Canu, S., Vera, P., & Ruan, S. (2021). Latent correlation representation learning for brain tumor segmentation with missing MRI modalities. *IEEE Transactions on Image Processing*, *30*, 4263-4274.

[27] Wang, B., Yang, J., Peng, H., Ai, J., An, L., Yang, B., ... & Ma, L. (2021). Brain tumor segmentation via multi-modalities interactive feature learning. *Frontiers in Medicine*, 341.

[28] Chen, X., Liew, J. H., Xiong, W., Chui, C. K., & Ong, S. H. (2018). Focus, segment and erase: an efficient network for multi-label brain tumor segmentation. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 654-669).

[29] Wang, G., Li, W., Ourselin, S., & Vercauteren, T. (2019). Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation. *Frontiers in computational neuroscience*, *13*, 56.
[30] Jiang, Z., Ding, C., Liu, M., & Tao, D. (2019, October). Two-stage cascaded u-net: 1st place solution to brats challenge 2019 segmentation task. In *International MICCAI brainlesion workshop* (pp. 231-241). Springer, Cham.
[31] Zhou, C., Ding, C., Wang, X., Lu, Z., & Tao, D. (2020). One-pass multi-task networks with cross-task guided attention for brain tumor segmentation. *IEEE Transactions on Image Processing*, *29*, 4516-4529.
[32] Cirillo, M. D., Abramian, D., & Eklund, A. (2020, October). Vox2Vox: 3D-GAN for brain tumour segmentation. In *International MICCAI Brainlesion Workshop* (pp. 274-284). Springer, Cham.
[33] https://www.med.upenn.edu/sbia/brats2018/evaluation.html.
[34] https://www.med.upenn.edu/cbica/brats2019/evaluation.html.
[35] https://www.med.upenn.edu/cbica/brats2019/data.html
[36] Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017, September). Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *International workshop on machine learning in medical imaging* (pp. 379-387). Springer, Cham.
[37] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
[38] He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision* (pp. 1026-1034).
[39] Huang, H., Yang, G., Zhang, W., Xu, X., Yang, W., Jiang, W., & Lai, X. (2021). A deep multi-task learning framework for brain tumor segmentation. *Frontiers in Oncology*, 2095.
[40] Liu, Z., Tong, L., Chen, L., Zhou, F., Jiang, Z., Zhang, Q., ... & Zhou, H. (2021). Canet: Context aware network for brain glioma segmentation. *IEEE Transactions on Medical Imaging*, *40*(7), 1763-1777.
[41] Sun, J., Peng, Y., Guo, Y., & Li, D. (2021). Segmentation of the multimodal brain tumor image used the multi-pathway architecture method based on 3D FCN. *Neurocomputing*, *423*, 34-45.
[42] Rehman, M. U., Cho, S., Kim, J., & Chong, K. T. (2021). Brainseg-net: Brain tumor mr image segmentation via enhanced encoder–decoder network. *Diagnostics*, *11*(2), 169.
[43] Zhang, D., Huang, G., Zhang, Q., Han, J., Han, J., Wang, Y., & Yu, Y. (2020). Exploring task structure for brain tumor segmentation from multi-modality MR images. *IEEE Transactions on Image Processing*, *29*, 9032-9043.
[44] Di Ieva, A., Russo, C., Liu, S., Jian, A., Bai, M. Y., Qian, Y., & Magnussen, J. S. (2021). Application of deep learning for automatic segmentation of brain tumors on magnetic resonance imaging: a heuristic approach in the clinical scenario. *Neuroradiology*, *63*(8), 1253-1262.
[45] Ali, M. J., Raza, B., & Shahid, A. R. (2021). Multi-level kronecker convolutional neural network (ML-KCNN) for glioma segmentation from multi-modal MRI volumetric data. *Journal of Digital Imaging*, *34*(4), 905-921.